**Data Analysis Reports**

***Steps taken into consideration***

|  |  |
| --- | --- |
| Sr.no | Task |
| 1. | Missing value analysis |
| 2. | Numerical variables analysis (distribution) |
| 3. | Outliers analysis |
| 4. | Categorical variables |
| 5. | Cardinality in categorical variables |
| 6. | Relationship between variables and target |
|  |  |

***Missing Value Analysis***

Variables with missing values:

['LotFrontage', 'Alley', 'MasVnrType', 'MasVnrArea', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinType2', 'Electrical', 'FireplaceQu', 'GarageType', 'GarageYrBlt', 'GarageFinish', 'GarageQual', 'GarageCond', 'PoolQC', 'Fence', 'MiscFeature']

Percentage of missing values in above variables:

LotFrontage 0.177397

Alley 0.937671

MasVnrType 0.005479

MasVnrArea 0.005479

BsmtQual 0.025342

BsmtCond 0.025342

BsmtExposure 0.026027

BsmtFinType1 0.025342

BsmtFinType2 0.026027

Electrical 0.000685

FireplaceQu 0.472603

GarageType 0.055479

GarageYrBlt 0.055479

GarageFinish 0.055479

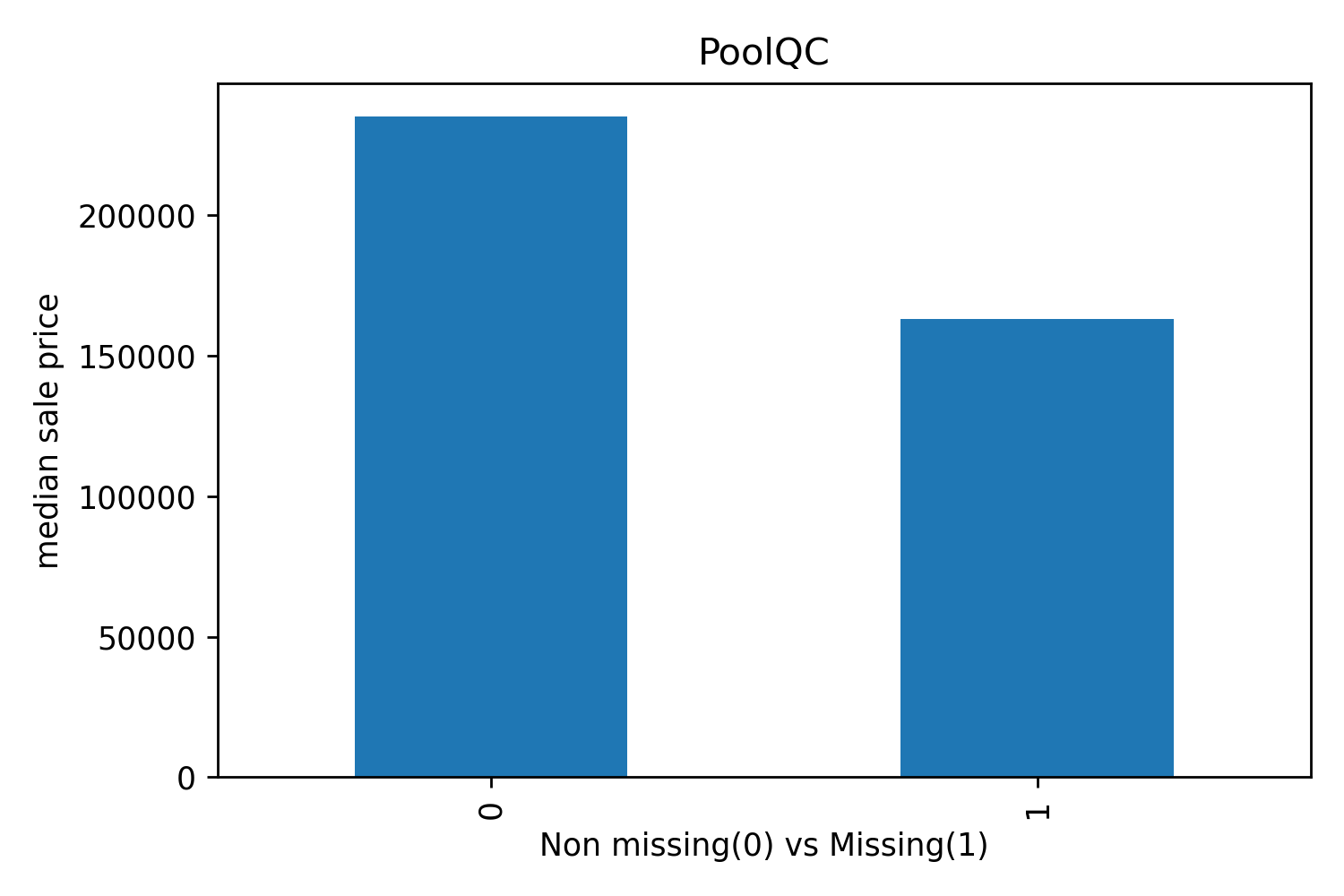
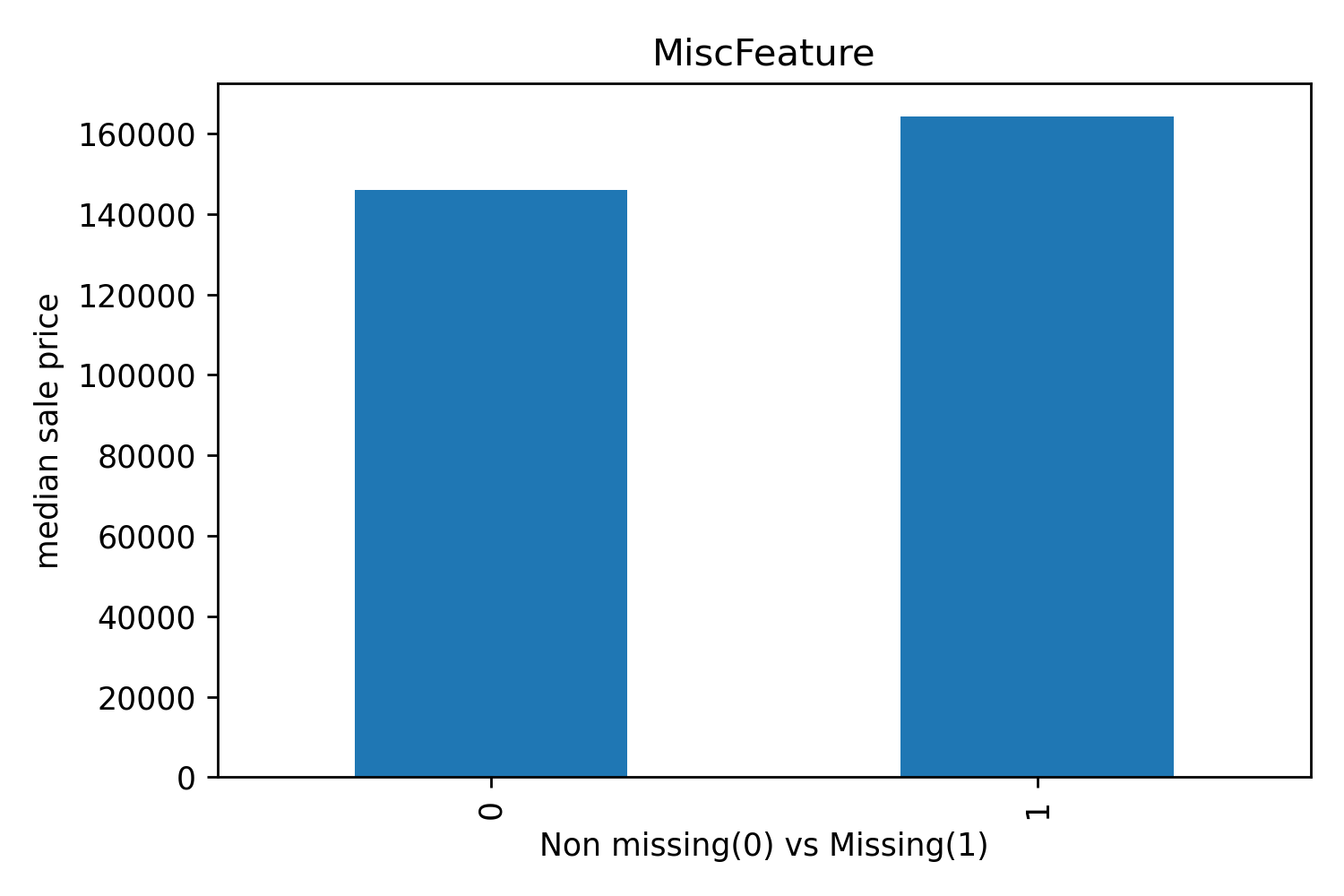
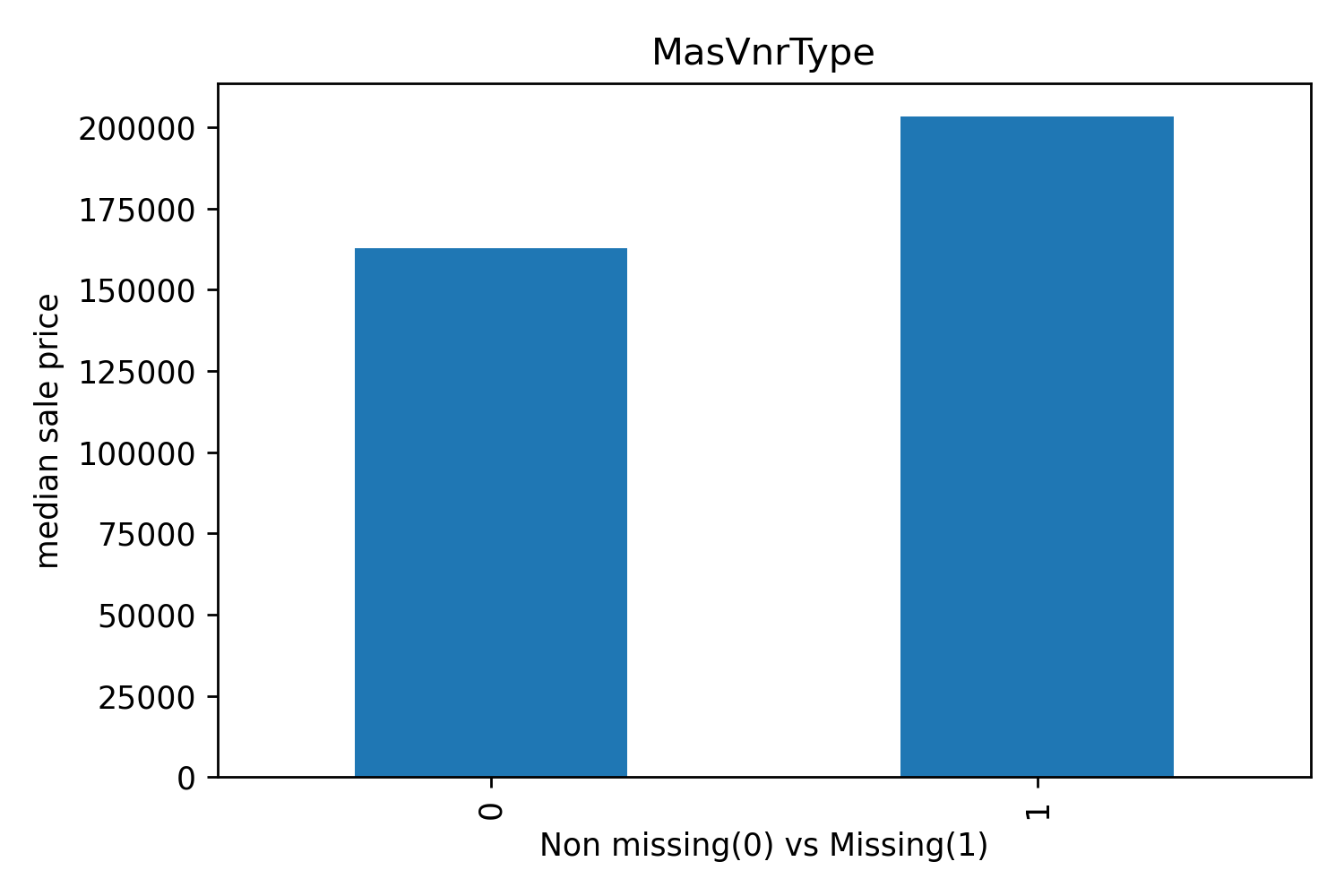
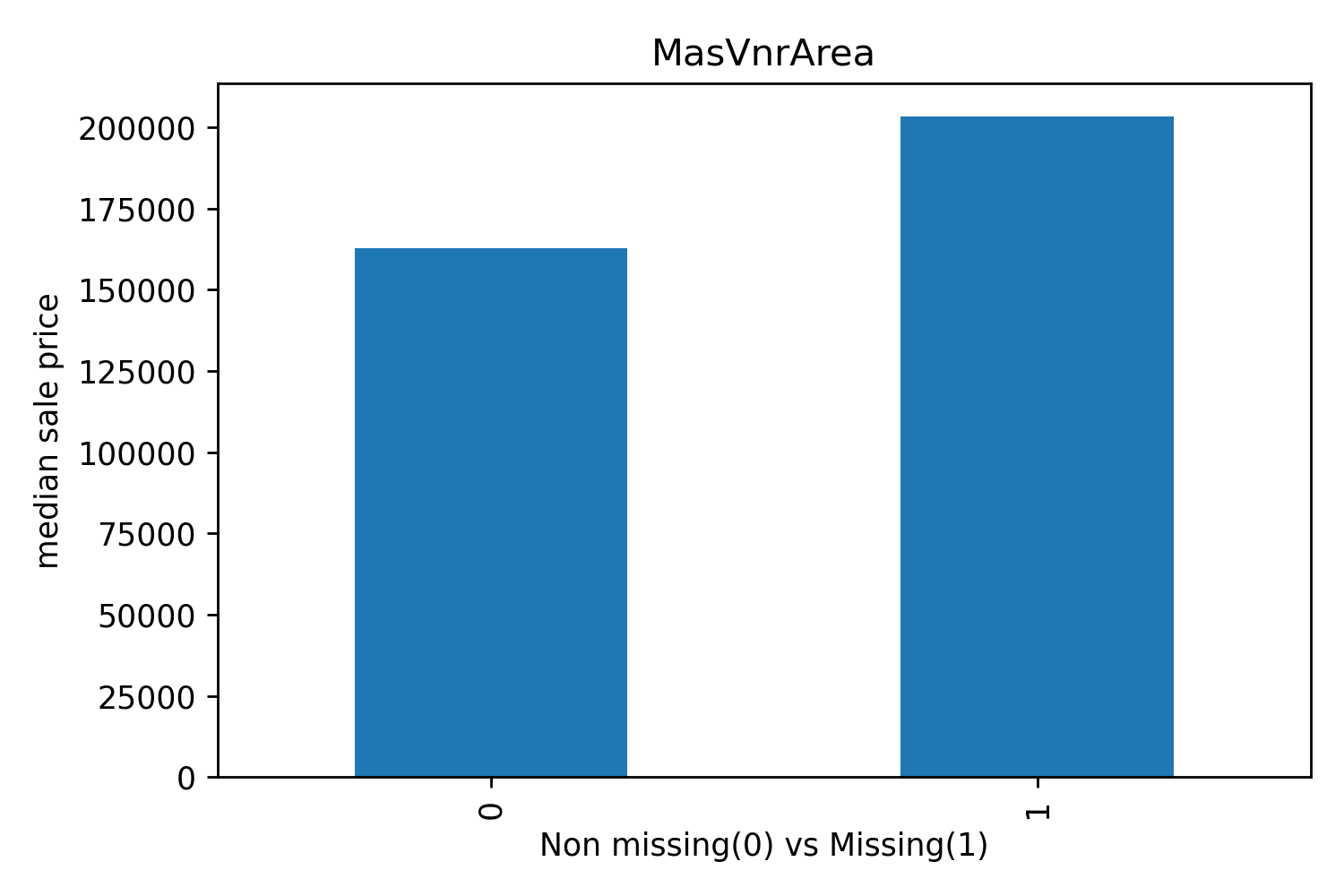
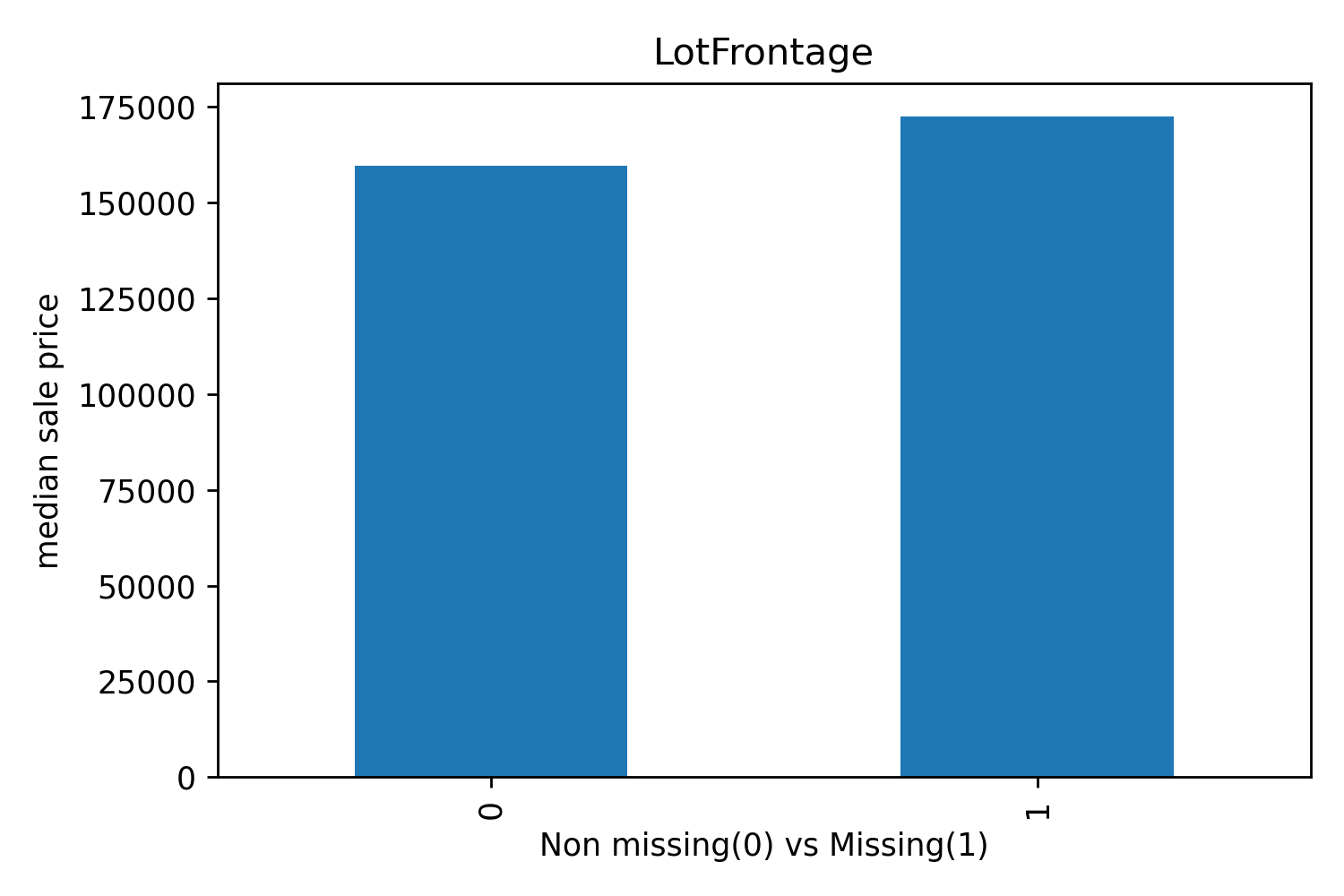
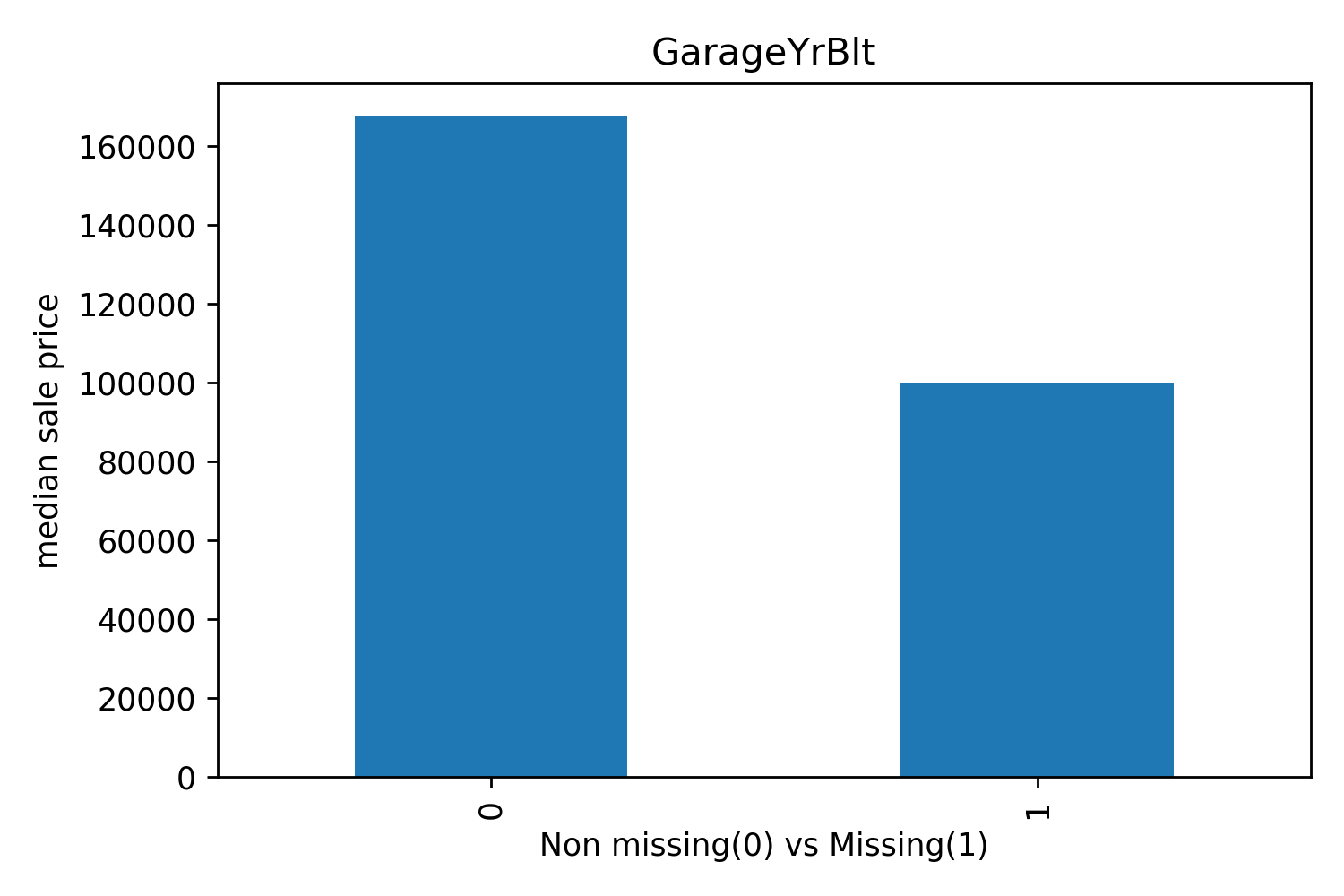
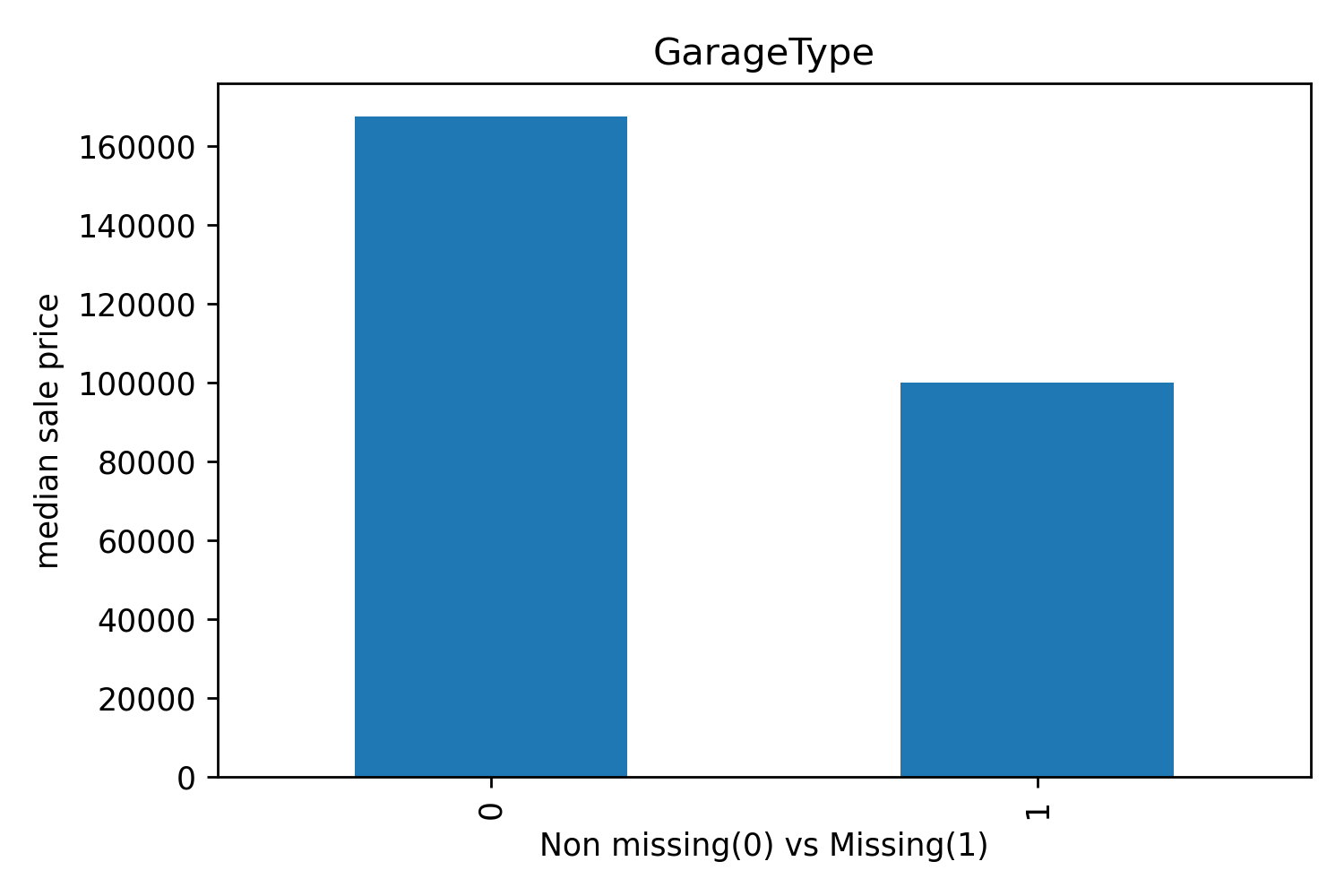
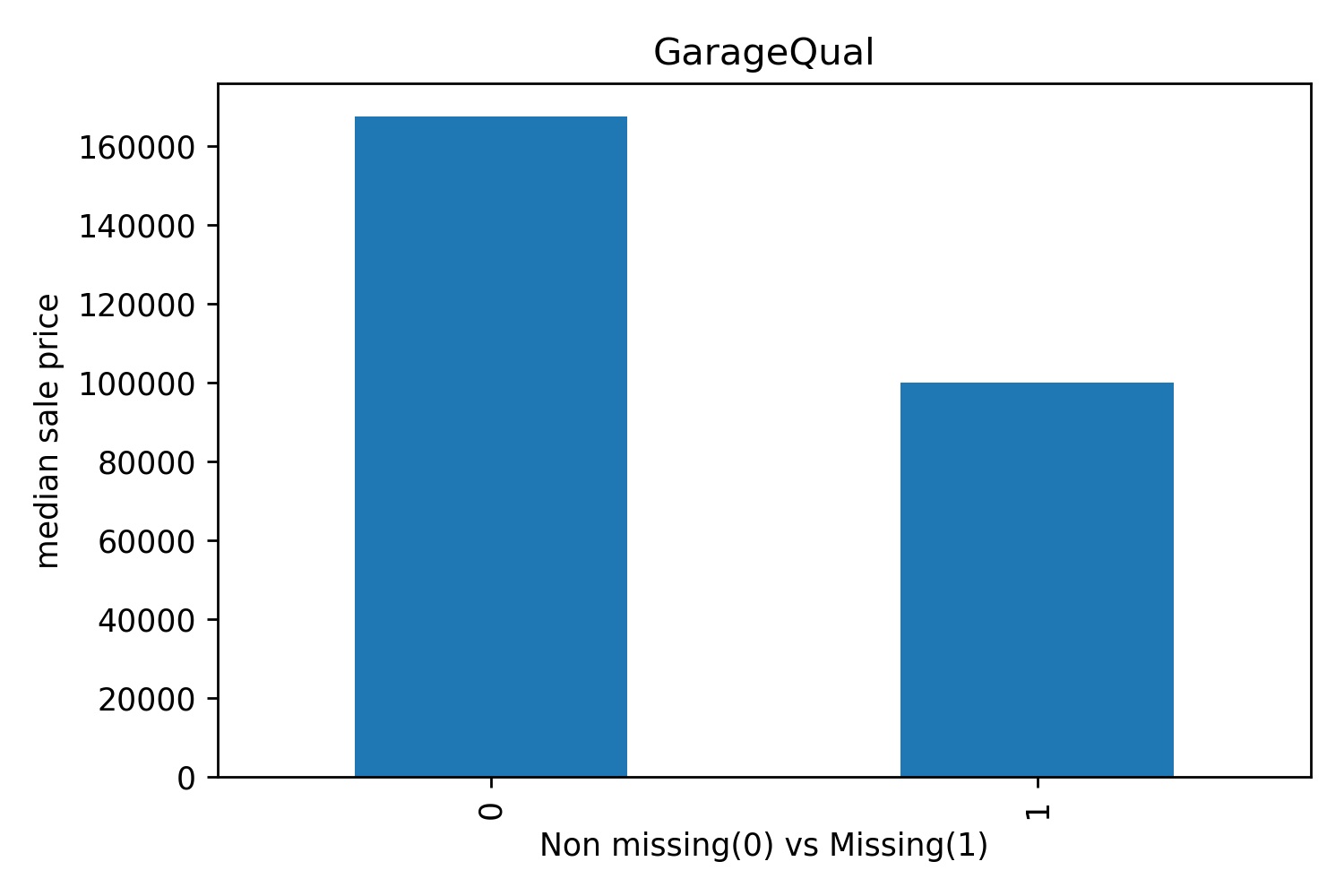
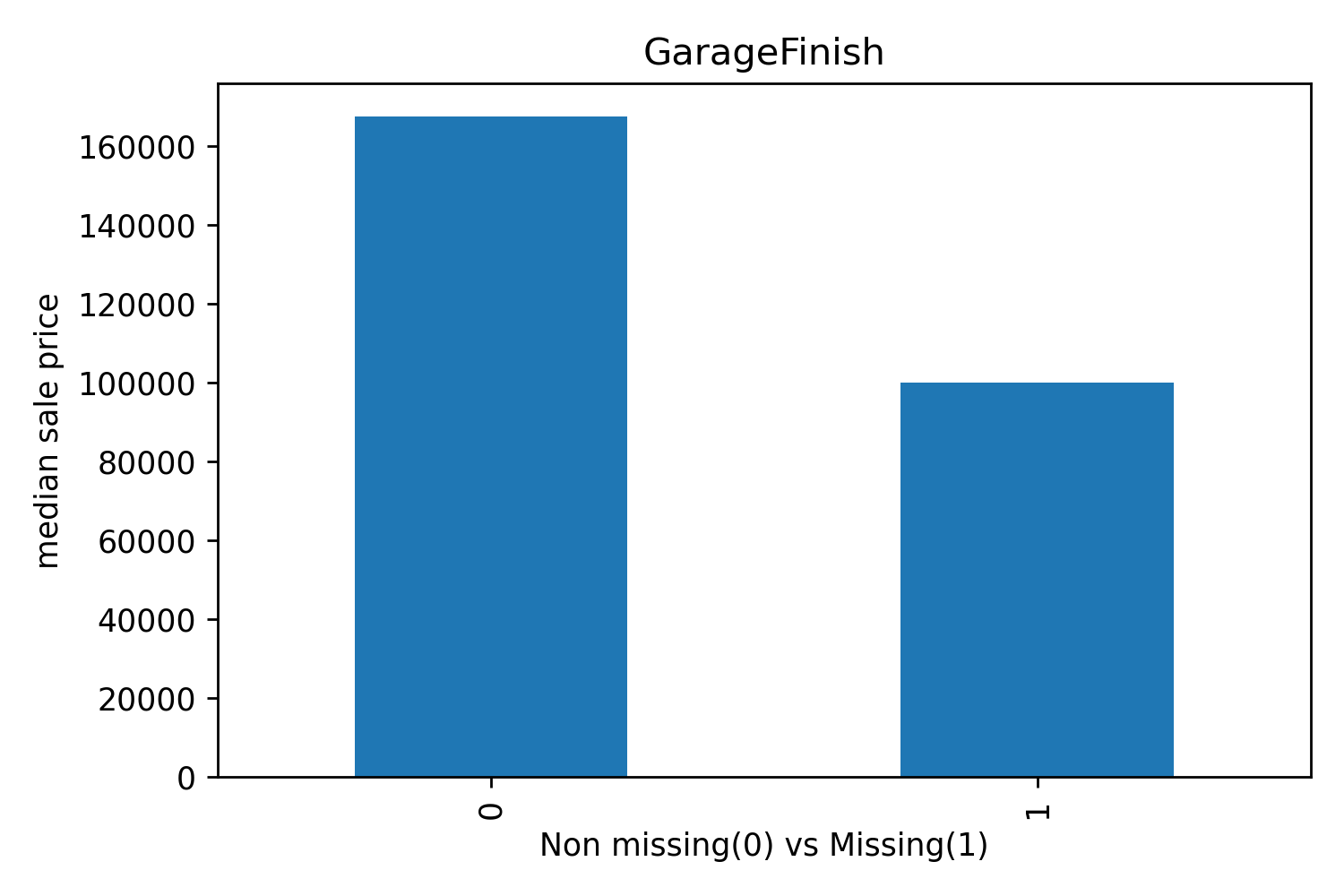
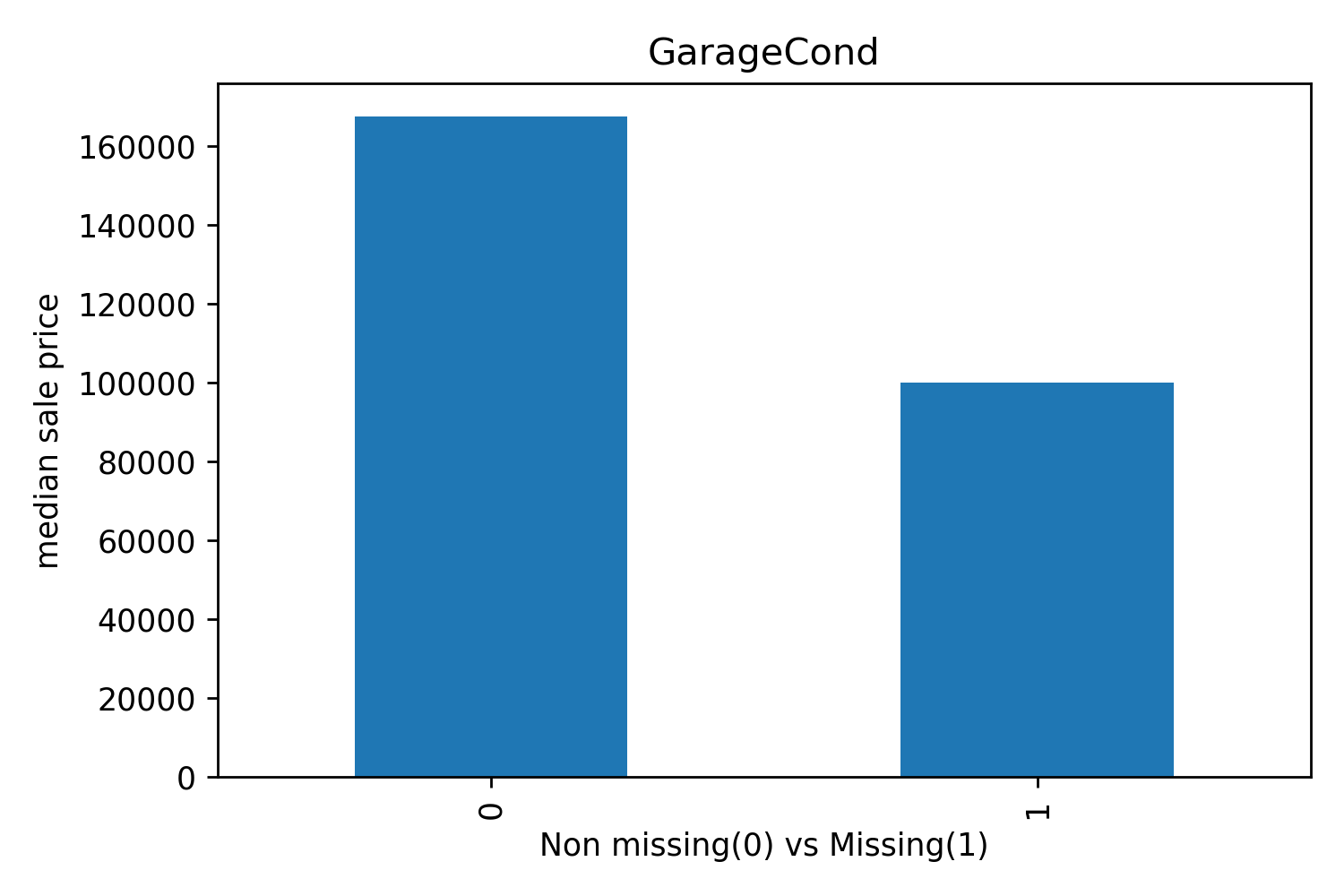
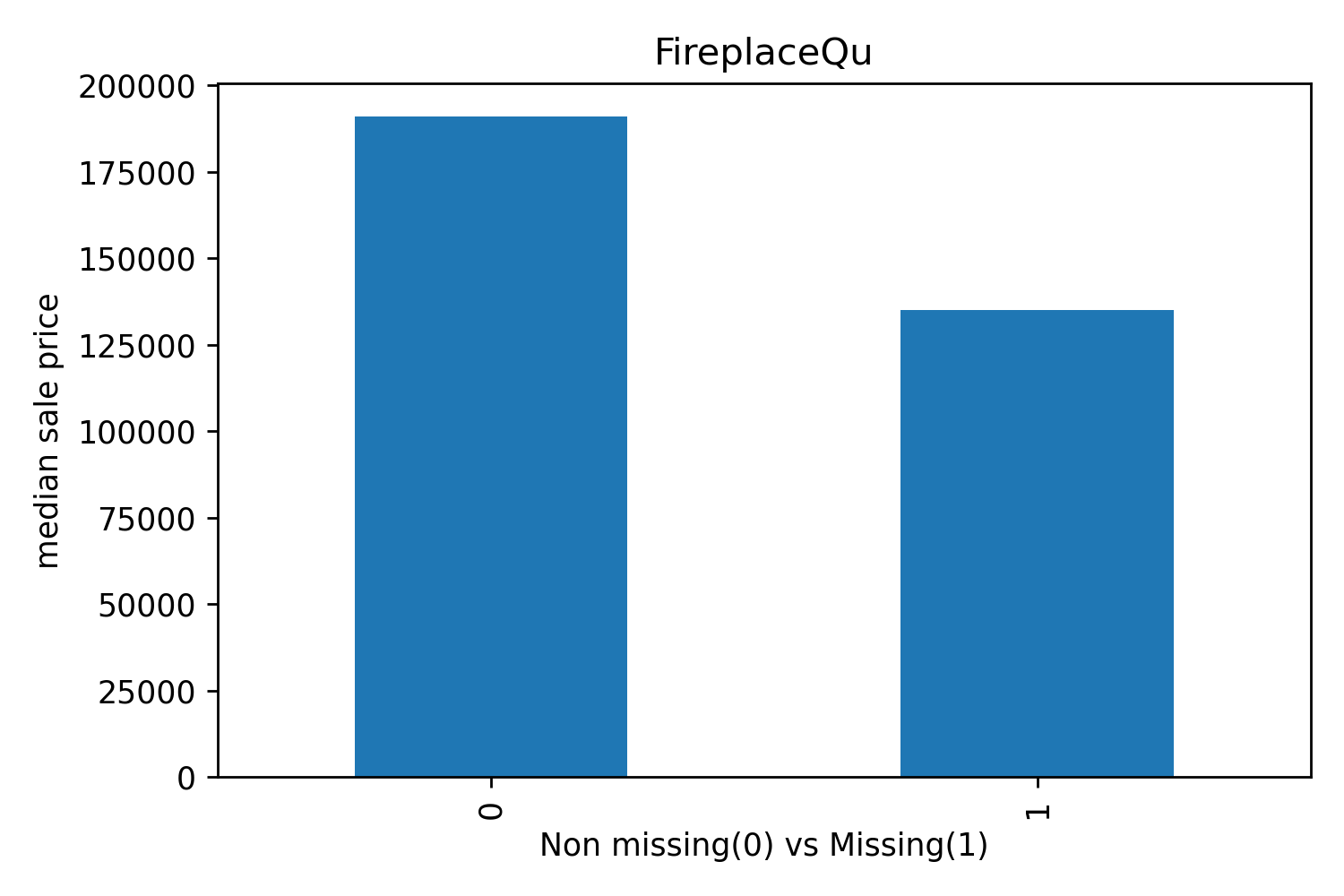
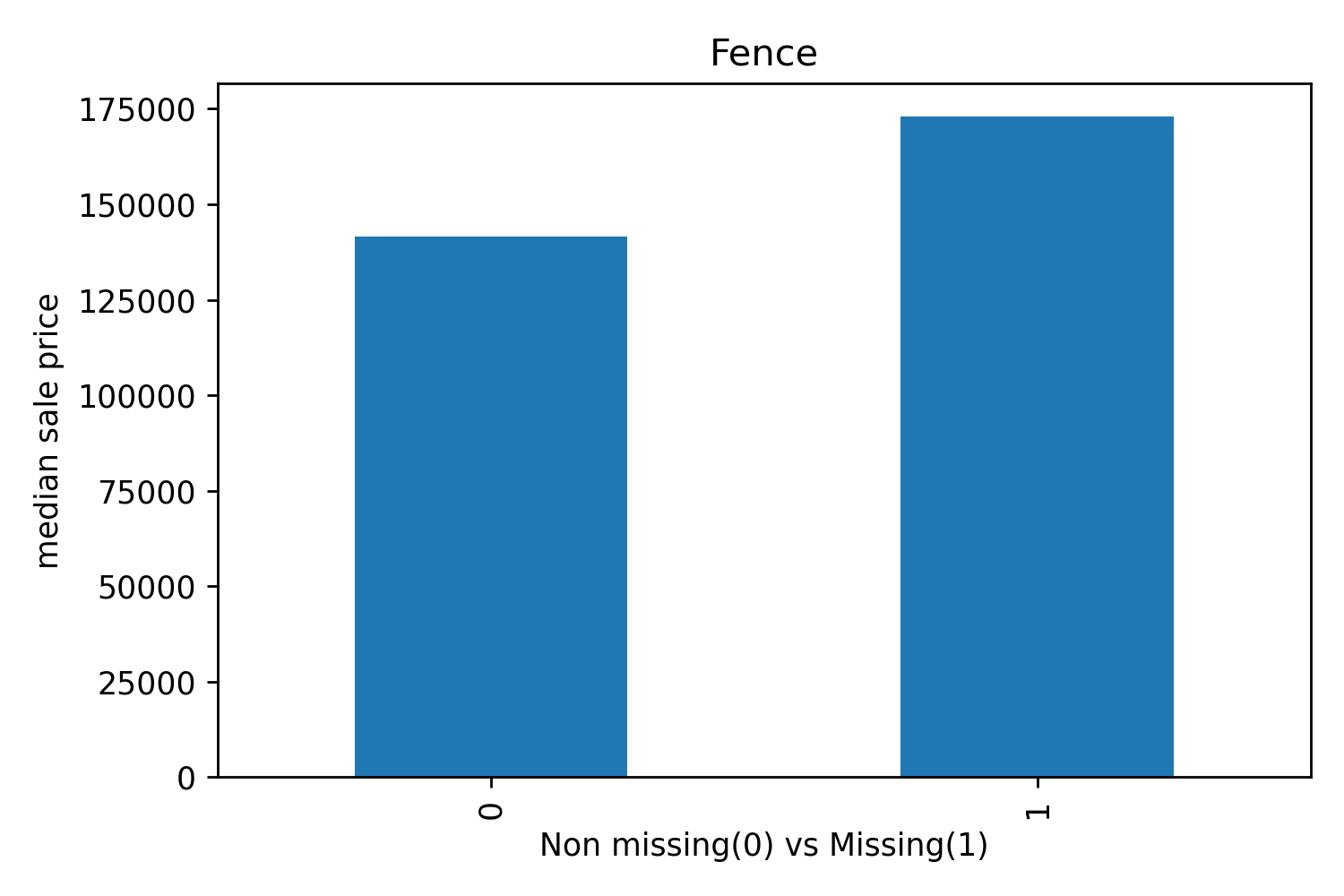
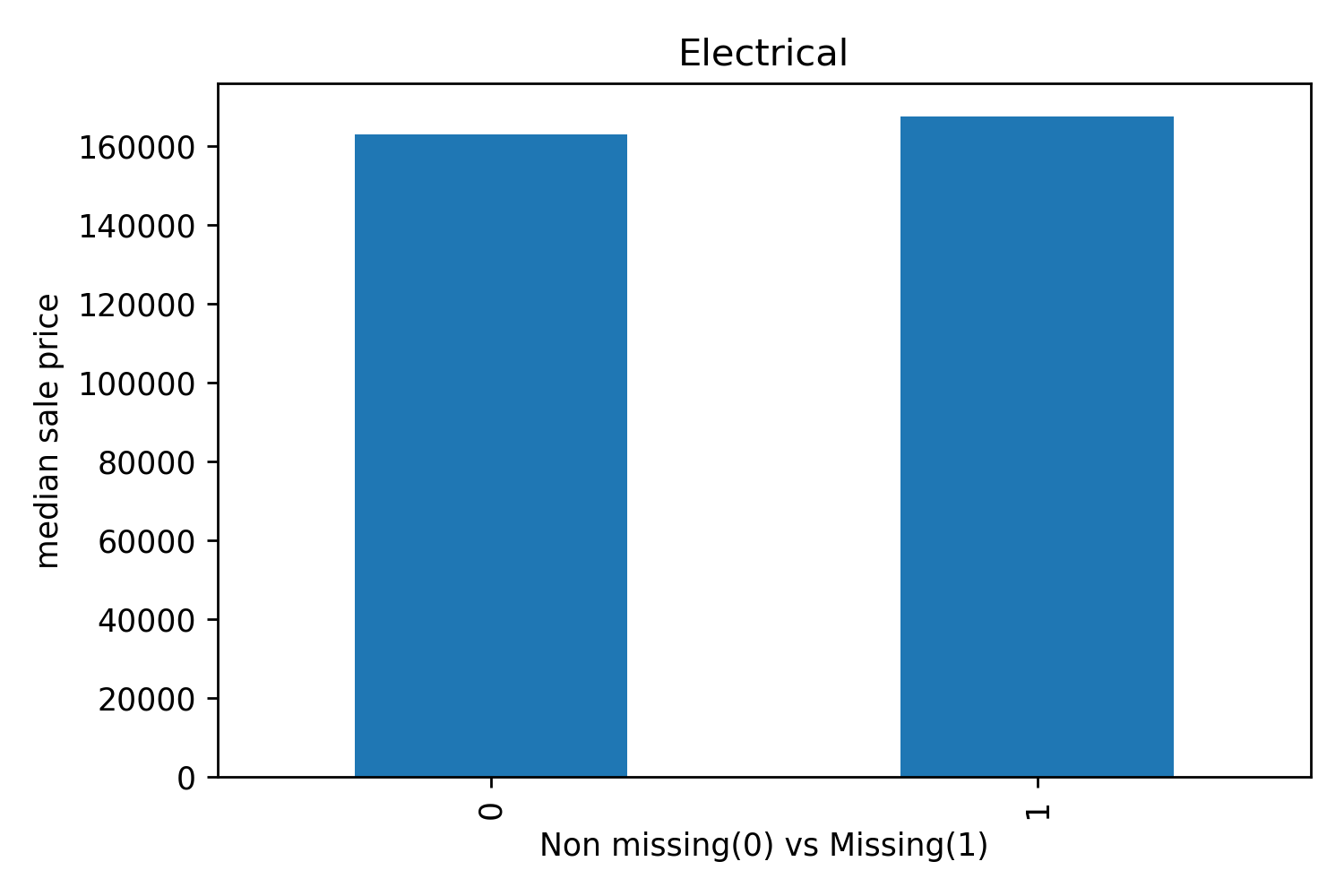
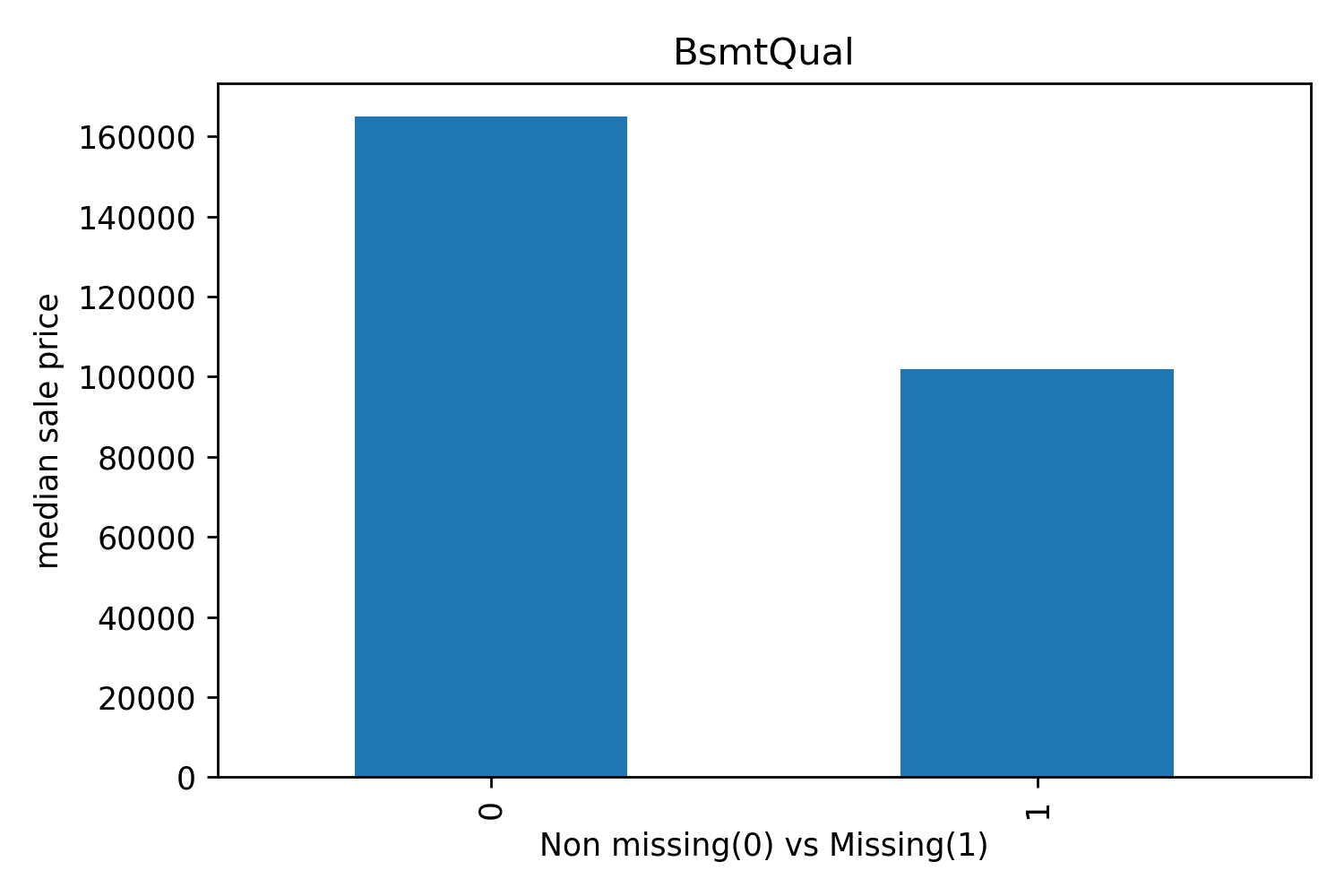
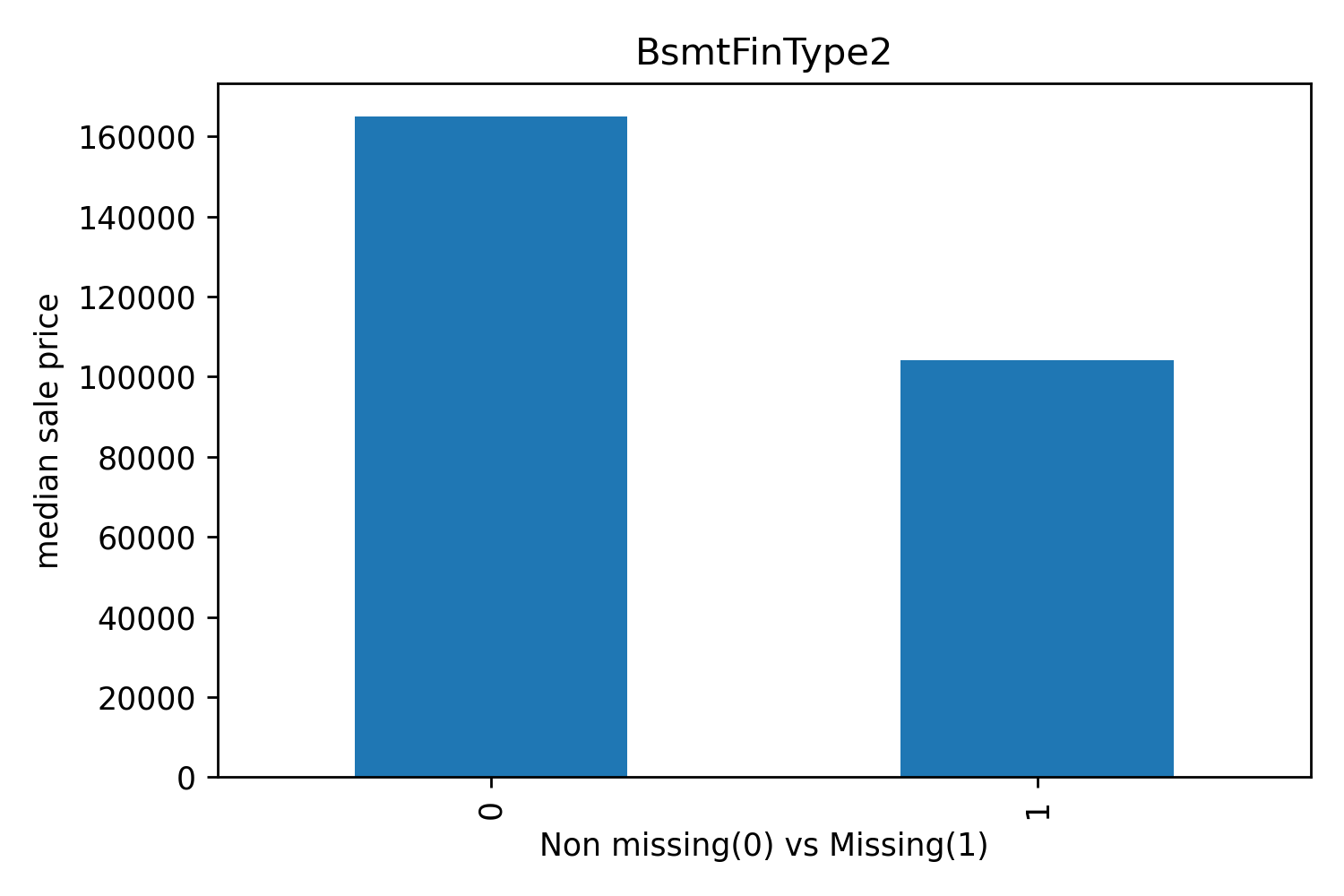
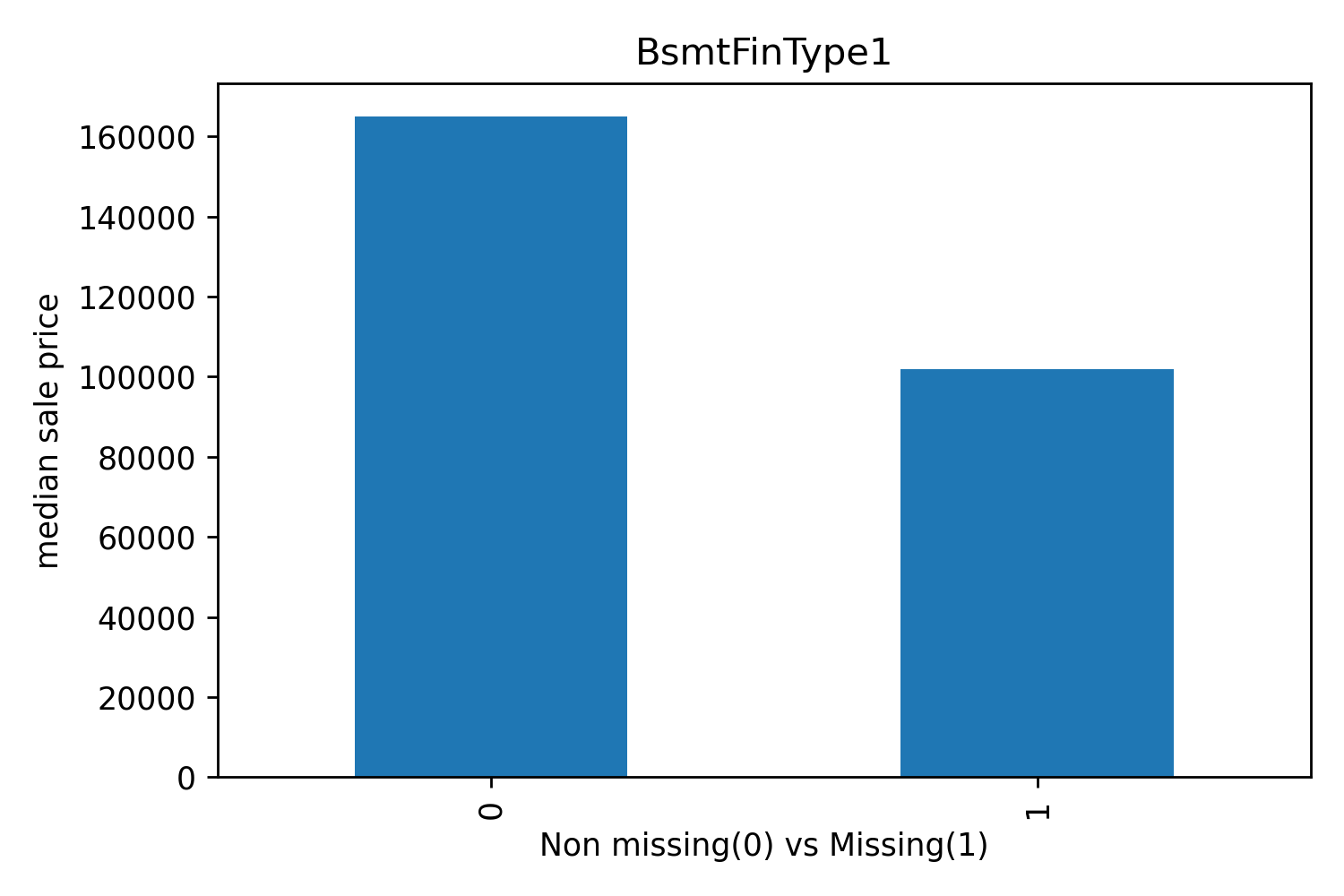
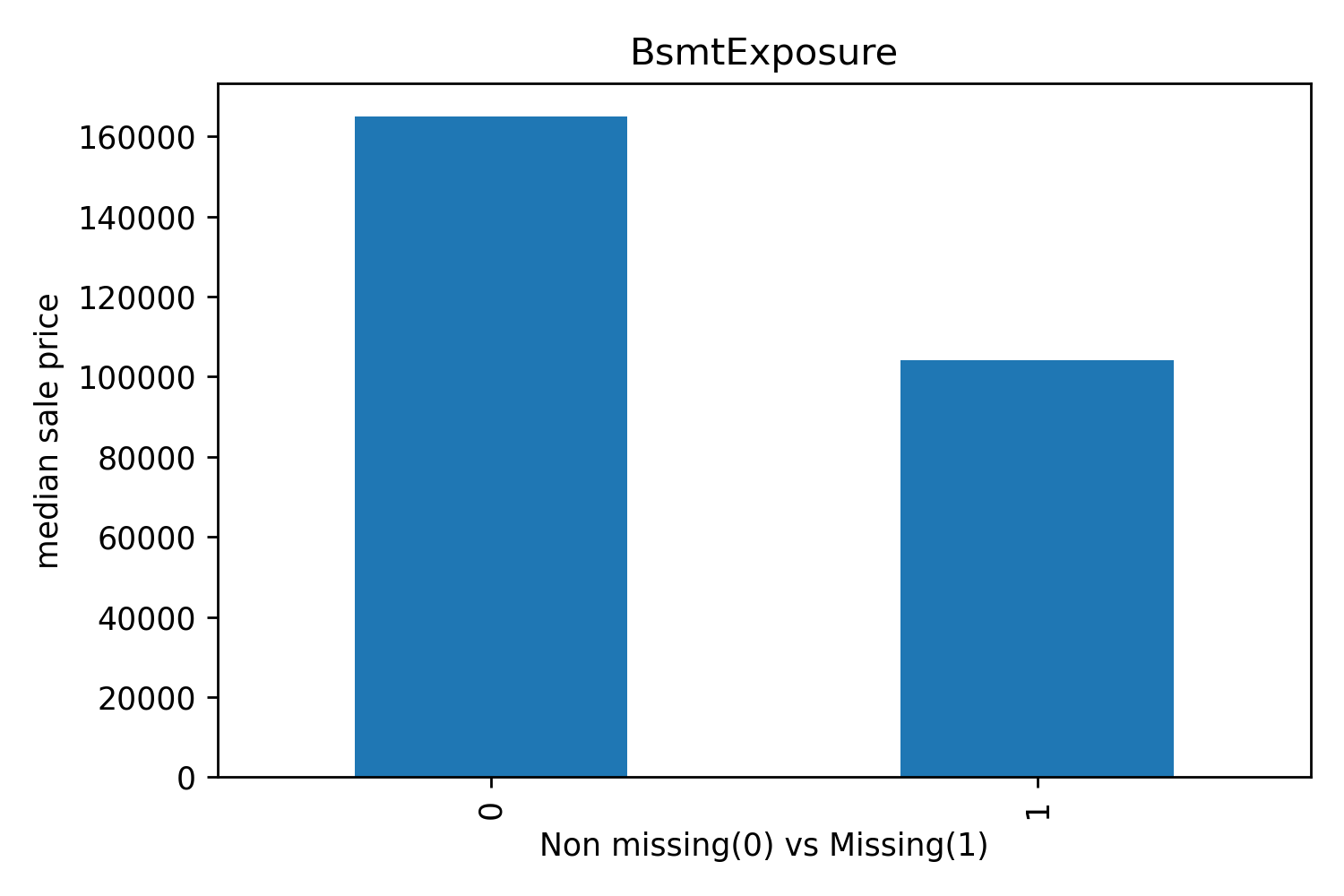
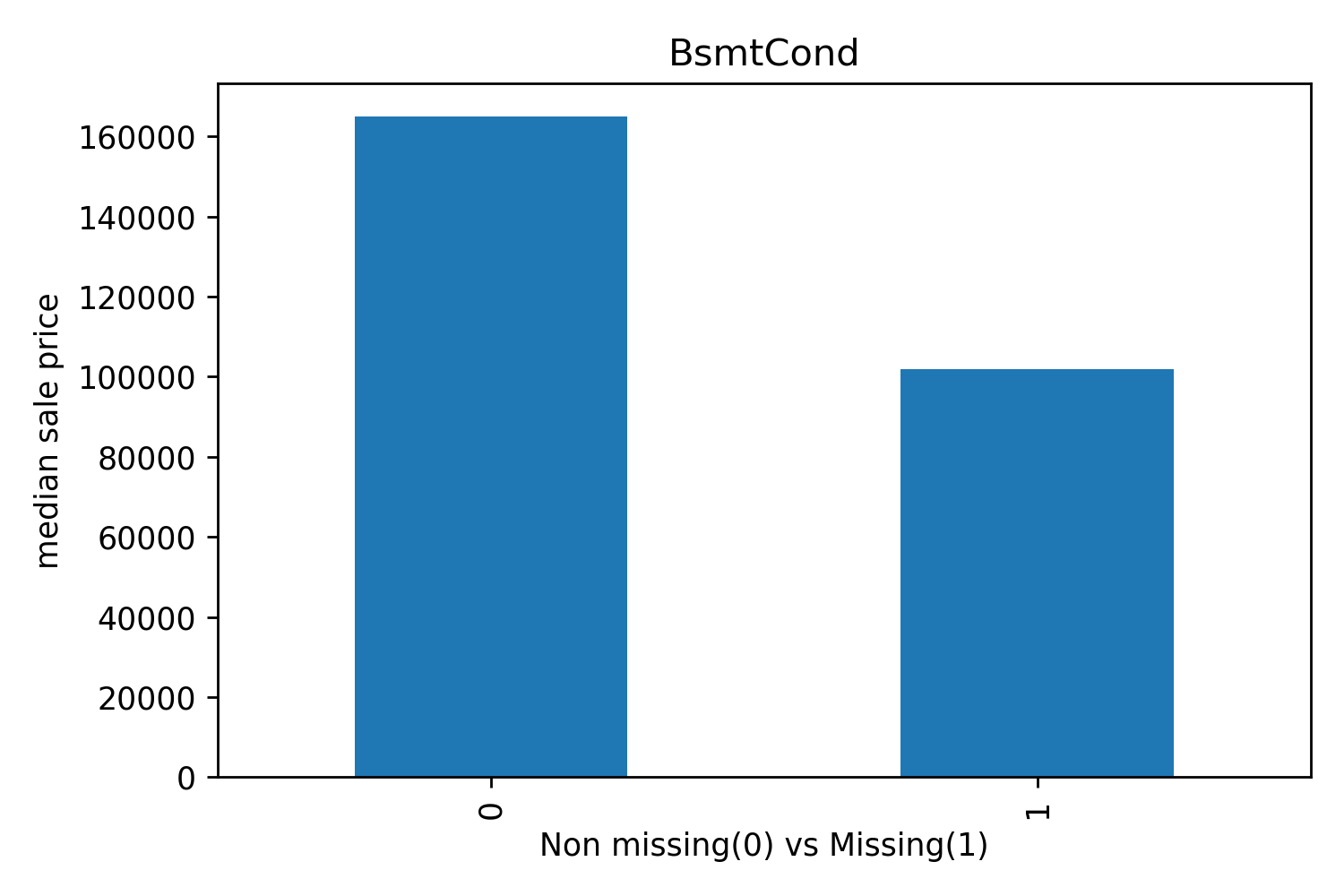
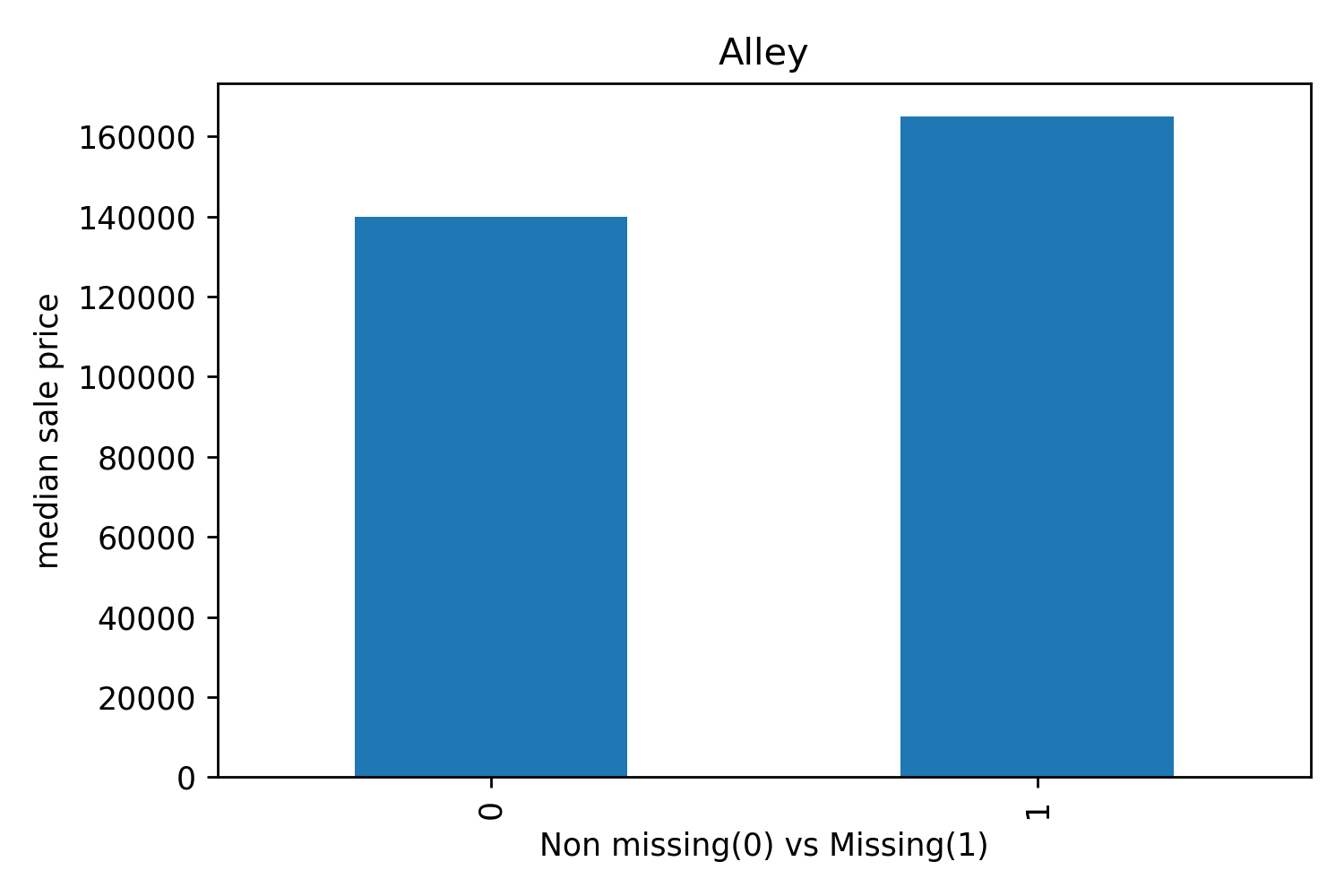
GarageQual 0.055479

GarageCond 0.055479

PoolQC 0.995205

Fence 0.807534

MiscFeature 0.963014

Impact of missing values on target:

Conclusion: The median house price is much smaller when the value in the dataset is missing